APPENDIX B

METHOD FOR CORRECTING RELATIVE RISK

FOR SMOKER MISCLASSIFICATION

APPENDIX B. METHOD FOR CORRECTING RELATIVE RISK FOR SMOKER MISCLASSIFICATION

B.1. INTRODUCTION

The purpose of this appendix is to present the details of the method used in Section 5.2.2. to correct observed passive smoking relative risks for the systematic upward bias caused by misclassification of some smokers as never-smokers. The method used is that proposed by A. J. Wells and W. F. Stewart (Wells, 1990) with minor modifications, including an adjustment for passive smoking risk to smokers. This appendix covers the following: the principles of the method (Section B.2); how the method differs from those previously used by the National Research Council and P. N. Lee (Section B.3); the data used to calculate the misclassification factors and other parameters (Section B.4); the mathematical model used to calculate the corrected relative risks (Section B.5); and a numerical example to show how the method is applied in a practical case (Section B.6). The results show that the bias due to smoker misclassification is highly unlikely to be responsible for the increased risks observed in the passive smoking lung cancer epidemiology studies. Evidence is also presented suggesting that the true downward corrections for smoker misclassification bias may be even smaller than those developed below and used in Section 5.2.2. While some of the rates presented below are subject to variability and argument, attempts are made to provide reasonable estimates and a defensible methodology.

There is considerable literature on this topic and a history of controversy regarding the magnitude of the bias and whether it may explain the observed increase in lung cancer mortality due to ETS exposure. The NRC report on the health effects of passive smoking (NRC, 1986) delves into this topic in considerable detail. It concludes that bias is likely; further, it estimates an adjustment for the summary relative risk from the combined results for all ETS studies. The NRC report further concludes that smoker misclassification does not account for the observed passive smoking risk. On the other hand, in various publications Lee (1987b, 1988, 1990, 1991a) has claimed that the smoker misclassification bias is large enough to explain most or all of the observed passive smoking lung cancer risk.

Approaches to estimation of misclassification bias have used mathematical modeling with parameters estimated from a variety of sources that have not always been consistent. The procedure described below attempts to rectify some previous sources of misunderstanding on this topic and utilizes the extensive data sources now available to improve parameter estimates and tailor refinements to individual populations.

B.2. PRINCIPLES OF THE WELLS-STEWART METHOD

The Wells-Stewart method is based on the following principles, the nature and need for which have largely become apparent from the chronological evolution and disparate approaches to and results for this problem.

The parameters are:

a. Since the passive smoking epidemiology is essentially concerned entirely with self-reported never-smokers, it is necessary to limit the misclassifieds to those who said they never smoked, not simply to nonusers, because the latter would include a substantial proportion of self-reported former smokers.

b. Use one minus sensitivity or its close relative, false negatives (misclassified smokers) divided by observed positives (self-reported smokers) as the vehicle for transferring misclassification data from cotinine and discordant answer studies to the passive smoking studies. Sensitivity is the term used to describe the fraction correctly classified as exposed, namely, true positives divided by true positives plus false negatives, but since we are assuming that the true positives and the observed positives are the same (no misclassification of never-smokers as smokers), sensitivity in this case becomes observed positives divided by observed positives plus false negatives. Then one minus sensitivity becomes false negatives divided by observed positives plus false negatives. Ignoring the false negatives in the denominator introduces negligible error. In any case, do not use specificity (true negatives divided by true negatives plus false positives) or any parameter that uses as its denominator true or observed negatives (self-reported never-smokers). The reason is that sensitivity is affected much less by smoker prevalence than parameters based on observed negatives.

c. Calculate a correction for each epidemiologic study separately using a misclassified smoker relative risk and a proportion of smokers among subjects and spouses that is characteristic of the timeframe and locale of each study. Use data from the study itself or from another study with the same target population, if possible.

d. Use only female data to correct misclassification of female subjects.

For the mathematical model, calculate the corrected risk directly--that is, do not first calculate a bias assuming no passive risk and then divide the observed risk by that bias to get a corrected risk.

Subjects found to be misclassified as nonsmokers are categorized according to their self-reported smoking status--former or current. Misclassified current smokers are further classified as "regular" or "occasional," according to observed cotinine levels. "Regular" means the cotinine level is above 30% of the self-reported smoker mean; "occasional" applies to the range 10% to 30%. Cotinine levels are not informative for misclassified former smokers, who tend to be long-term abstainers (10+ years, according to Lee [1987b] and Wald et al. [1986]). The two studies with detailed cotinine levels on female current smokers (Lee, 1986 and Haddow et al., 1986, in Table B-1) indicate that about 10% of the current smokers are occasionals.

B.3. DIFFERENCES FROM EARLIER WORK

The Wells-Stewart method differs from the method used by the NRC (1986), which is also described by Wald et al. (1986), in that the NRC method failed to separate the misclassified smokers into regular, occasional, and ex-smokers, and failed to account for the effect of smoker misclassification on active smoker risk. The NRC made an overall correction to the aggregated passive relative risk using United Kingdom (U.K.) smoking prevalence and risk rather than making the corrections study-by-study with appropriate smoking prevalences and risk for each study's time and locale, and it mixed male data with female data in arriving at misclassification factors. Their calculated bias of 1.34/1.25 = 1.07, or 7%, for the combined worldwide studies is substantially higher than the 2% overall bias that would result if the biases in Table 5-7 were aggregated. The discrepancy is largely due to NRC's use of U.K. parameters for all of the studies regardless of locale, plus some overestimation of the impact of misclassified occasional and ex-smokers.

Lee's methods have evolved over the years in three stages. In Lee (1987b, 1988), he improved on the NRC method in that he divided the misclassified smokers into ex-smokers and current regular and occasional smokers, and he corrected the smoker risk for misclassification. However, all of the five principles listed above were violated to some degree, resulting in about a twelvefold overestimation of the bias (Wells, 1992). The Lee (1990) paper correctly limits misclassifieds to never-smokers, relates misclassified smokers to smokers, not to never-smokers, and treats each study separately, but still mixes male input data with female data for use in calculating bias for females. Furthermore, his mathematical model still relies on the assumption of a passive smoking relative risk of 1.00 (no risk), an assumption that fails at passive risks above about 1.3 and overstates those biases. In addition, Lee (1990) has changed from separating the misclassified smokers into three groups in favor of the (less useful) overall category of "ever-smokers." Most recently, Lee (1991a) presented a more complex mathematical model that includes a term for passive risk, but the method still has the other shortcomings noted for Lee (1990). A comparison of the most recent Lee bias estimates with those in Table 5-7 is shown in Table B-2 for the five U.S. studies with the greatest statistical weight. When Lee's inputs are used with the Wells-Stewart mathematical model, the calculated biases are, if anything, somewhat larger than when using Lee's most recent model. Therefore, the difference between Lee's most recent

Table B-1. Observed ratios of occasional smokers to current smokers (based on cotinine studies)

	Females			Both sexes[1]		
Study	Occl.[2]	Current	Occl./current	Occl.	Current	Occl./current
Lee (1986)	4	72	0.056	12	176	0.068
Coultas et al. (1988)				59	278	0.212
Haddow et al. (1986)	10	64	0.156			
Feyerabend et al. (1982)[3]				7	82	0.085
Jarvis (1987)				12	90	0.133
Pojer (1984)				25	187	0.134
Wald et al. (1984)				13	131	0.099
Overall	14	136	0.103	128	944	0.136

[1]The "both sexes" data are shown to indicate that the female value of 10.3% is not unduly high.
[2]Occasional smokers are defined as persons who have cotinine levels in body fluids that are between 10% and 30% of the mean of all self-reported current smokers.
[3]The Feyerabend et al. (1982) data are for nicotine.

estimates of bias and those shown in Table 5-7 are in practical terms due almost entirely to differences in input parameters. The input parameters we have chosen are developed in the next section, and comparisons with the Lee parameter estimates are shown as footnotes to Table B-2.

B.4. PARAMETER ESTIMATES

The key input in these calculations is the proportion of misclassified regular current smokers who claim they have never smoked. Our definition of misclassified regular current smokers, first suggested by Lee (1987b), produces a mean cotinine level approximately equal to that of all self-reported current smokers. Detailed data from three large cotinine studies have been assembled for use herein with the cooperation of their principal investigators (Coultas, Cummings, and Pierce in Table B-3). The data identify individual nonsmokers with cotinine values greater than 10% of the mean for self-reported smokers, by sex and self-reported smoking status (never or former). Data on nonusers are also available from several other studies (the lower

Table B-2. Examples, using five U.S. studies, of differences in smoker misclassification bias between EPA estimates and those of P.N. Lee regarding passive smoking relative risks for females

Study	% of U.S. weight	Lee (1991a) model[1] Lee (1991a) input parameters			Wells-Stewart model Lee (1991a) input[2] parameters			Wells-Stewart model EPA input parameters (Table 5-8)		
		RR_o	RR_c	Bias	RR_o	RR_c	Bias	RR_o	RR_c	Bias
FONT	35	1.32	1.18	1.11	1.32	1.13[3]	1.16	1.29	1.28[3]	1.01
GARF (Coh)	25	1.17	1.02	1.14	1.17	1.02[4]	1.14	1.17	1.16[4]	1.01
GARF	15	1.23	1.10	1.12	1.23	1.08[5]	1.14	1.31	1.27[5]	1.03
JANE	10	0.75	0.62	1.21	0.75	0.61[6]	1.24	0.86	0.79[6]	1.09
CORR	3	2.07	1.84	1.12	2.07	1.70[7]	1.22	2.07	1.89[7]	1.10

Note: Calculated bias is very sensitive to three key factors, high values of which will drive the bias up; namely, fraction of observed never-smokers misclassified, female active smoker relative risk, and female smoking prevalence. Lee's inputs are higher than EPA's, as indicated in footnotes 2 to 7 below. RR_o = observed passive risk. RR_c = passive risk corrected for smoker misclassification bias. Bias = RR_o/RR_c.

[1]Additive model, Lee's Table 3. His additive model was chosen because it is similar to our additive model for passive smoking effects on smokers.

[2]EPA's misclassification factors developed in Section B.4., namely, 1.09% of current regular smokers, 24.2% of current occasional smokers, and 11.7% of ex-smokers, when weighted for their respective prevalence and relative risk, are equivalent to about 1.5% of average self-reported ever-smokers. EPA used these rates for all studies except FONT, which is a special case. Lee used 2.0% of self-reported ever-smokers for all studies.

[3]Lee used 49% ever-smokers vs. our 43% based on the case age distribution. Our misclassification rates for current smokers, m_2 (4.3%) and m_3 (0%), were developed as noted in Section B-4, except that 2 out of 3.5 expected misclassified occasional smokers had been eliminated by cotinine tests, leaving 1.5/35 = 0.043 for m_2 in this study. For m_1, we assumed that it was the same percentage (41%) of 0.117 as 10% was of 24.2% for m_2.

(continued on the following page)

Table B-2. (continued)

[4]A female smoker risk of 3.58 (U.S. DHHS, 1986) and smoker prevalence of 22% (Hammond, 1966) for age distribution of cases. Lee used 8.0 and 49%.
[5]EPA estimates a smoker risk of 6 and a smoking prevalence of 34% for the time period 1971-81 vs. Lee's values of 8.0 and 49%.
[6]The main difference is in the assumed smoker misclassification rate, but Lee's assumption of 49% smoking prevalence vs. 42% assumed by EPA increases the bias estimate from 1.09 to 1.15.
[7]Lee assumed 58% smoking prevalence vs. 47%, which EPA got from the paper itself. Lee assumed a lower smoker risk (9.5) vs. EPA's 12.4; the effect of this was offset by Lee's higher misclassification rates.

Table B-3. Misclassification of female current smokers

Study	Cotinine level[1]	Self-reported smoking status number: Number — Never	Former	Current
Coultas et al. (1988)[2]	10-30	7	3	
	30+	5	8	
	All	387	79	184
Cummings (1990)[3]	10-30	0	1	
	30+	2	0	
	All	225	143	116
Pierce et al. (1987)[4]	10-30	9	4	
	30+	3	3	
	All	232	79	167
Subtotal	10-30	16	8	(67% never)
	30+	10	11	(48% never)
	All	844	301	467
Lee (1986)[5]	10-30	3	2	
	30+	3	2	
	All	333	125	256
Haddow et al. (1986)[5]	10-30	1	1	
	30+	0	1	
	All	174	58	64
Haddow et al. (1988)[5]	10-30	15	7	
	30+	1	1	
	All	1,128	380	503
Riboli (1991)[5,6]--U.S.[7]	10-30	1	0	
	30+	0	0	
	All	224	81	143
Riboli (1991)[5,6]--East Asia[8]	10-30	1	1	
	30+	1	0	
	All	325	25	77
Riboli (1991)[5,6]--Greece[9]	10-30	0	0	
	30+	0	0	
	All	96	5	15
Total	10-30	37	19	
	30+	15	15	
	All	3,124	975	1,525
Proportion misclassified[10]	10-30%	24.2%	12.4%	
	30+	1.09%	1.09%	

(continued on the following page)

Table B-3. (continued)

[1]Cotinine levels are in units of percentages of the mean of self-reported smokers for each study; 30+% are defined as current regular smokers, 10-30% are occasional smokers.
[2]Dr. Coultas kindly provided the individual cotinine values for females ages 18+ that were used in Table 3 of their paper. The totals differ slightly from the totals in the paper.
[3]Dr. Cummings kindly provided the cotinine levels for the six misclassified current smokers, three males and three females. As noted in the paper, current smokers were recruited during only the first half of the study. Therefore, the total equivalent current smokers were estimated from the current smoker/never-smoker ratio from national statistics.
[4]Individual cotinine levels for the misclassifieds by gender are from a personal communication from Petra Macaskill, who now has the basic data for this study.
[5]For Lee (1986), Haddow et al. (1986, 1988), and Riboli (1991), no breakdown was given between "Never" and "Former." An estimate was made based on the subtotal distribution. The number of smokers had to be estimated in some cases. The mean for self-reported smokers for Haddow et al. (1988) was very low, at 145 ng/mL, because the women were pregnant.
[6]Personal communication--individual country data from Riboli et al. (1990).
[7]New Orleans, Los Angeles, and Honolulu.
[8]China (Shanghai), Hong Kong, and Japan (Sendai).
[9]Athens.
[10]The observed current smokers are assumed to be 90% regular (1,372) and 10% occasional (153) smokers. For regular smokers, misclassification as never-smokers is 15/1,372 = 1.09% of observed current regulars or 15/(1,372 + 15 + 15) = 1.07% of true current regulars. For occasional smokers, misclassification is 37/153 = 24.2% of observed current occasionals or 37/(153 + 37 + 19) = 17.7% of true current occasionals. For current smokers misclassified as former smokers, the factors are 15/1,372 = 1.09% for observed and 15/1,402 = 1.07% for true regular smokers, and 19/153 = 12.4% for observed and 19/209 = 9.1% for true occasionals.

portion of Table B-3). The proportions of misclassified smokers who would have said "never" versus "former" are estimated using the proportions observed in the first three studies. Data sets not differentiating outcomes by sex have not been used. Also, the large 1987 study by Haddow and colleagues has not been used for this purpose on the advice of one of the authors (personal communication from G.J. Knight). This study of the effect of current smoking on birthweight relied on the cotinine data to distinguish smokers from nonsmokers. The questionnaire data were not collected in a manner that could be equated to the care that would be taken in either their or others' passive smoking studies.

The number of self-reported never- and former smokers with sufficiently high cotinine levels to be reclassified as current smokers is shown by study in Table B-3. As described above, those with cotinine levels in the 10-30% range are considered to be occasional smokers, whereas those above 30% are treated as regular smokers. If it is assumed (Table B-1) that 1,372 (90%) of 1,525 self-reported current smokers are regular smokers, leaving 153 (10%) as occasionals, then the percentage of current regular smokers misclassified as never-smokers totalled over all studies in Table B-3 is 15/1,372 or 1.09%. The percentage is almost the same if the number of true, i.e., self-reported plus misclassified current regular, smokers is used. For the occasional smokers only, the misclassification rate is much higher, about 24% (18%) of observed (true) occasional smokers. It is possible, however, that the subjects classified as occasional smokers based on cotinine levels in the range 10-30% may contain some true never-smokers that are just highly exposed to passive smoke.

The cutoff points used, namely, 30% of the self-reported current smoker mean cotinine level to distinguish misclassified regular smokers from occasional smokers and 10% of the self-reported current smoker mean cotinine level to distinguish occasional smokers from current nonsmokers, were chosen originally by Lee (1987b). They are justified as follows: the actual cotinine levels of the 15 misclassified current smokers in the Never column of Table B-3 whose levels exceeded 30% of the mean cotinine level for self-reported current smokers in each study were divided by the mean smoker cotinine level for that study. These values were then averaged for each study, and a mean for all studies was obtained by weighting each study's mean by the number of smokers in that study. The overall mean cotinine level for the misclassified smokers was 94% of the mean for all of the self-reported smokers because the misclassifieds tended to concentrate near the bottom of the 30%+ range. A cutoff of 35% could be justified since the misclassifieds' mean cotinine level was 99% of the mean for the self-reported smokers, but we chose to continue with 30% to be conservative.

The cutoff between the current nonsmokers and the occasional smokers must be somewhat arbitrary because there is an overlap between heavily ETS-exposed nonsmokers and very light current smokers. Authors who have tried to eliminate all possible smokers from their cohorts have used lower cutoff points. For example, Coultas et al. (1988), Cummings (1990), and Haddow et al. (1988), who were trying to eliminate smokers, used cutoffs between 7% and 8%. However, Pierce et al. (1987) and Lee (1986), who, as we are, were trying to distinguish smokers from nonsmokers, used higher cutoffs, 16% and 9%, respectively. The mean of the percentages (calculated as above for the misclassified current regular smokers) that the misclassified occasional smokers' cotinine levels bear to the mean of the self-reported current smokers is 16% for the seven studies in Table B-3. This is lower than the midpoint of the 10-30% range, again because the individual values concentrate at the lower end of the range. If we had used a 5% cutoff instead of 10%, the misclassification rate for occasional smokers would have been increased from 24% to about 40%, but the average of the percentages of current self-reported mean cotinine levels for the misclassified occasional smokers would have dropped from 16% to 13%. This in turn would reduce the estimated smokers' relative risk for this group, and the overall effect on the corrected risk of never-smokers would be negligible.

The studies in Table B-4 provide data on discordant answers, i.e., reported never-smokers who have called themselves smokers on one or more previous occasions. Based on those data, the estimated percentage of former smokers misclassified as never-smokers is 11.7% (10.8%) of the observed (true) number of former smokers. As mentioned previously, evidence suggests (Wald et al., 1986; Lee, 1987b) that most former smokers misclassified as never-smokers have been nonsmokers for an extended period, such as 10+ years, and may have been light smokers on average. Accordingly, we have used a weighted average of the data of Alderson et al. (1985), Lubin et al. (1984), and Garfinkel and Stellman (1988) for 10+ year abstainers to estimate misclassified former smoker relative risk, namely, an excess risk that is 9% of current self-reported smoker excess risk.

Some confusion and misleading conclusions on smoker misclassification have resulted from the practice of expressing the number of smokers misclassified as never-smokers as a percentage of the total number of (either true or observed) never-smokers, rather than as a percentage of the number of smokers. That leads to a higher expected percentage of smokers misclassified as never- smokers among cases than controls because lung cancer cases are much more likely to have been smokers than never-smokers. Some people (Lee, 1988) have interpreted a higher percentage of observed never-smokers later found to be misclassified smokers among the cases as evidence that smokers with lung cancer are more apt to claim falsely to be never-smokers than persons without cancer. That conclusion, however, appears to be an artifact of treating the misclassification rate as a percentage of the number of never-smokers rather than as a percentage of the number of smokers. The study data summarized in Table B-5 do not support that conclusion. If anything, they are more supportive of the conclusion that ever-smokers in lung cancer studies may be less likely to misrepresent themselves as never-smokers than members of the general public who are questioned in community surveys. The 1.0% average misclassification rate shown in Table B-5 for the lung cancer cases suggests that estimates such as the 5.7% from the general population studies (Table B-5) or the equivalent of 3.9% of ever-smokers (Table B-4) that we have used may be much too high. Further corroboration that the misclassification rates from the community studies are too high relative to those in the epidemiologic studies is found in the recent lung cancer case-control study by Fontham et al. (1991), which specifically included in its design a screening by urinary cotinine levels to eliminate current smokers from both cases and controls. After eliminating possible smokers among the self-reported never-smokers by the usual epidemiologic questionnaire and medical records review techniques, the investigators found by cotinine measurements that only two probable occasional smokers and no probable regular smokers were left among the 239 never-smoking lung cancer cases for which cotinine measurements were made. Using the procedures herein and assuming 43% ever-smokers among controls and an ever-smoker

Table B-4. Misclassification of female former smokers reported as never-smokers based on discordant answers

Study	Locale	Former smokers (FS)[1]	Ever-smokers (ES)[1]	Reported never-smokers who reported earlier that they had smoked[1] N	Percentage of ES	Percentage of FS
Kabat and Wynder (1984)[2]	U.S.					
Controls		109	319	0	0.0	0.0
Cases		222	652	7	1.1	3.2
Machlin et al. (1989)	U.S.	194	687	52	7.6	26.8
Krall et al. (1989)[3]	Mass.	11	30	1	3.3	9.1
Britten (1988)[4]	U.K.	320	878	38	4.3	11.9
Lee (1987b)	U.K.	85	243	13	5.5	15.3
Akiba et al. (1986)	Japan	8	38	0	0.0	0.0
Overall[5]		949	2847	111	3.9	11.7

[1]Number of former smokers and ever-smokers had to be estimated in some cases.
[2]Dr. Kabat (personal communication) advised that of 13 misclassifieds, 8 were females, 1 of whom used snuff.
[3]Krall data are based on 20-year recall.
[4]Britten data include only those persons who said they never smoked but actually had smoked regularly one or more cigarettes per day.
[5]For former smokers, misclassification as never-smokers would appear to be 111/949 = 11.7% of observed former smokers or 111/(949 + 111) = 10.5% of true former smokers, but from Table B-3 16 + 15/(16 + 15 + 975) = 3.08% of former smokers are really current smokers, so the 949 + 111 = 1,060 should be reduced by 3.08% to 1,027 as the number of true former smokers. Then 111/1,027 = 10.81%, based on true former smokers.

Table B-5. Misclassification of female lung cancer cases

Source	Number of ever-smokers	Number misclassified
CHAN Chan et al. (1979)[1]	12	1
KABA Kabat and Wynder (1984)[2]	652	7
AKIB Akiba et al. (1986)	38	0
PERS Pershagen et al. (1987)	179	2
HUMB Humble et al. (1987)[3]	223	1
Total	1,104	11 (1.0%)
General population[4]	1,838	104 (5.7%)

[1]Chan sampled five Type I and II never-smokers, one of whom was said by a relative to have smoked a few hand-wrapped cigarettes for a year at age 71. The ratio of smoking to nonsmoking cases for Types I and II was 44/19, which, multiplied by 5, leads to 12 estimated ever-smokers.
[2]Dr. Kabat (personal communication) advised that of 13 misclassifieds, 8 were females, 1 of whom used snuff.
[3]Of the four misclassifieds found, Dr. Humble (personal communication) has advised that most if not all were males. We have assumed one female.
[4]The general population data are taken from the four nonlung cancer cohorts in Table B-4, namely, Machlin et al. (1989), Krall et al. (1989), Britten (1988), and Lee (1987b).

relative risk of 8, which translates to 10 for misclassified current regular smokers, 2.44 for misclassified occasionals, and 1.81 for misclassified ex-smokers, there would have been 1,363 smoker cases, consisting of 1,328 current smokers and 35 occasional smokers to go along with 420 never-smoking cases. It is seen that a misclassification rate of 0/1,328 = 0.00% for regular smokers is well below the 1.09% that we have used from the surveys in Table B-3. For occasionals, there would be 20 cases to go along with 239 never-smoking cases, yielding a misclassification rate of 2/20 = 10%, which is also well below the 24.2% for occasionals that we have used from Table B-3.

Another indication that the estimates based on community surveys may be too high comes from analysis of male data. The observed percentage of never-smokers is typically much lower for males (17% to 35%) than females (41% to 86%). To correct for smoker misclassification, we set up a table analogous to Table B-6 where the number of current and former smokers

Table B-6. Deletions from the "never" columns in Tables B-13 and B-16 and corrected elements

		Wife's smoking status					
Husband's smoking status		Former (1)	Occl. (2)	Regular (3)	Sum[1] (4)	Observed never (5)	Corrected never[2] (6)
Table B-13	Never	0.00679	0.00194	0.00081	0.00953	0.286	0.27647
(controls)	Ever	0.01275	0.00532	0.00219	0.02027	0.242	0.22173
Table B-16	Never	0.00198	0.00120	0.00217	0.00534	0.052	0.04666
(cases)	Ever	0.00770	0.00365	0.00604	0.01739	0.092	0.07461

[1](4) = (1) + (2) + (3)
[2](6) = (5) - (4)

misclassified as never-smokers are subtracted from the reported number of never-smokers. When the misclassification rates generated from community surveys are applied to the male data, the outcome is not credible--the number deleted for misclassification exceeds the total number of reported never-smokers in 3 of the 11 examples of which we are aware and drives the corrected relative risk well below unity in 4 more. This outcome indicates that the misclassification rates derived from the community surveys are too high. It is probable that the true smoker misclassification bias is on the order of one-fourth to one-half of the values shown in Table 5-7.

It has also been suggested (Lee, 1991b) that East Asian women misclassify themselves at much higher rates than Western women. The data from the International Agency for Research on Cancer (Riboli, personal communication) in Table B-3 do not support that claim, however, because the East Asia (Hong Kong, Japan, and China) misclassification rate for current regular smokers is 1/77 = 1.3%, which is not much different from the overall rate of 1.09%.

In conclusion, it would appear that the bias introduced by misclassification of smokers as never-smokers is not a serious problem. It probably increases observed excess relative risks on a worldwide basis by about 1% and for combined U.S. studies by about 3%.

B.5. MATHEMATICAL MODEL

The proportion of smokers, m_{h0}, misclassified as never-smokers is estimated separately for former smokers (m_{10}), occasional smokers (m_{20}), and regular smokers (m_{30}). Similarly, the proportion of current smokers, m_{h1}, misclassified as former smokers is estimated separately for occasional smokers (m_{21}) and regular smokers (m_{31}). These estimates are given in Tables B-3 and B-4. It is assumed that there is no misclassification of true never-smokers as current or former smokers or of true former smokers as current smokers. Also, these misclassification factors are used for all the studies unless otherwise noted. We suspect that misclassification rates probably vary from study to study. That variability, however, would tend to cancel out as the individual study results are combined.

Let c_{ijk} designate the observed distribution of controls (i = 0) and cases (i = 1) by their smoking status (j = 0,1,2,3) and the smoking status of their husbands (k = 0,1), as illustrated in Table B-7. Following the notational convention that a dot in the subscript position means summation on that subscript, then $c_{0\cdot\cdot} = c_{1\cdot\cdot} = 1$.

The observed c_{ijk}'s are corrected for misclassification of the wife's smoking status by first specifying a 4 × 4 matrix of distribution (Table B-8), where P_{hj} (h,j = 0,1,2,3) is the probability that a subject with true smoking status h will also be observed to have smoking status j. The subscripted notation is shown in Table B-8 for easy reference. $P_{\cdot\cdot}$ is equal to unity.

For passive smoking, we are interested only in correcting the c_{i0k} values that are for the observed never-smokers. It is assumed that the P_{hj}'s are the same for cases and controls (nondifferential misclassification). For given values of wife's subject status (i) and husband's smoking status (k), the correction when the wife's observed smoking status is "never" (j = 0) is:

$$C_{i0k} = c_{i0k} - \sum_{h=j=1}^{3} c_{ijk} \, (P_{h0}/P_{\cdot j}) \tag{B-1}$$

where C_{i0k} is the corrected form of the element c_{i0k}. Then the corrected passive risk, RR(c), becomes:

$$RR(c) = (C_{101} \times C_{000})/(C_{100} \times C_{001}) \tag{B-2}$$

The values of c_{0jk} in Table B-7 are from prevalence data in the study itself or from a related study, from concordance data, and from each study's data on the smoking prevalence of the never-smokers' husbands. If necessary, the number of former smokers can be estimated from the ever-smokers based on data from nine studies known to us where the percentage of both current smokers and former smokers is known (see Table B-9). These data indicate a time trend in nontraditional societies, from 20% former smokers relative to ever-smokers in 1960 to 45% in 1985; we estimate an 8-year lag for the traditional societies such as Hong Kong, China, Japan, and Greece, based on the data in Koo et al. (1983) and Sobue et al. (1990).

Table B-7. Notation for distribution of reported female lung cancer cases and controls by husband's smoking status

Wife's subject status (i)	Husband's smoking status (k)	Wife's observed smoking status (j): Never (j = 0)	Ex (j = 1)	Occl. (j = 2)	Reg. (j = 3)	Total
Control	Never (k = 0)	c_{000}	c_{010}	c_{020}	c_{030}	$c_{0\cdot 0}$
(i = 0)	Ever (k = 1)	c_{001}	c_{011}	c_{021}	c_{031}	$c_{0\cdot 1}$
	Total	$c_{00\cdot}$	$c_{01\cdot}$	$c_{02\cdot}$	$c_{03\cdot}$	$c_{0\cdot\cdot}$ (= 1)
Case	Never (k = 0)	c_{100}	c_{110}	c_{120}	c_{130}	$c_{1\cdot 0}$
(i = 1)	Ever (k = 1)	c_{101}	c_{111}	c_{121}	c_{131}	$c_{1\cdot 1}$
	Total	$c_{10\cdot}$	$c_{11\cdot}$	$c_{12\cdot}$	$c_{13\cdot}$	$c_{1\cdot\cdot}$ (= 1)

Table B-8. Notation for distribution of subjects by observed and true smoking status

Wife's observed smoking status (j)	Wife's true smoking status (h): Never (h = 0)	Former (h = 1)	Occl. (h = 2)	Reg. (h = 3)	Total
Never (j = 0)	P_{00}	P_{10}	P_{20}	P_{30}	$P_{\cdot 0}$
Former (j = 1)	P_{01}	P_{11}	P_{21}	P_{31}	$P_{\cdot 1}$
Occl. (j = 2)	P_{02}	P_{12}	P_{22}	P_{32}	$P_{\cdot 2}$
Reg. (j = 3)	P_{03}	P_{13}	P_{23}	P_{33}	$P_{\cdot 3}$
Total	$P_{0\cdot}$	$P_{1\cdot}$	$P_{2\cdot}$	$P_{3\cdot}$	$P_{\cdot\cdot}$(= 1)

Table B-9. Observed ratios of female former smokers to ever-smokers in the U.S., U.K., and Swedish studies: populations or controls (numbers or percentage)

Study	Time-frame	Never-smokers	Current smokers	Former smokers	Ever-smokers	Former/ever-smokers
Hammond (1966)[1]	1960	78.0%	17.6%	4.4%	22.0%	0.20
Buffler et al. (1984)[2]	1978	41%	38%	21%	59.0%	0.36
Wu et al. (1985)[2]	1980	92	73	55	128	0.43
Lee (1987b)[3]	1980	48.3%	33.6%	18.1%	51.7%	0.35
Brownson et al. (1987)[2]	1980	47	11	8	19	0.42
Britten (1988)[3]	1982	767	558	320	878	0.36
Humble et al. (1987)[2]	1982	162	63	48	111	0.43
Svensson et al. (1989)[2]	1984	120	53	36	89	0.40
Garfinkel and Stellman (1988)[1]	1982	58.9%	18.7%	22.4%	41.1%	0.54

Assumed ratios by years (nontraditional societies)[4]

Year	1960	1965	1970	1975	1980	1985
Ratio	0.20	0.25	0.30	0.35	0.40	0.45

[1]Using age distribution of never-smoking cases.
[2]Using age distribution of ever-smoking cases.
[3]Smoking status of general population.
[4]Traditional societies (Japan, Greece, China, Hong Kong) are estimated to lag these ratios by about 8 years, based on data in Koo et al. (1983) and Sobue et al. (1990). However, because the bias for the traditional societies is very low, changes in values of this parameter have little effect.

To calculate the individual elements, c_{0jk}, of Table B-7, it is necessary to establish concordance factors--that is, the cross products in 2 x 2 tables of smoking status of husbands and wives by smoking level of the wives. Using data from Sutton (1980), Lee (1987b), Akiba et al. (1986), and Hirayama (1984) and the detailed data in Lee (1987b) on never-smokers, current smokers, and former smokers, we have calculated that an appropriate average concordance factor for current smoking wives and ever-smoking husbands versus never-smoking wives and never-smoking husbands is 3.2; for ever-smoking wives and husbands versus never-smoking wives and husbands, it is 2.8, and for former smoking wives and ever-smoking husbands versus never-smoking wives and husbands, it is 2.2. These concordance factors can be expected to vary from study to study, but the effect of the variability should tend to cancel out as the studies are aggregated. The element c_{00}. and a quantity $s_0 = \sum_{j=1}^{3} c_{0j}$. are obtained from smoking prevalence data in the study itself, in a related study on the same cohort, or as a last resort from national statistics. If national statistics are used, care must be taken to use the rates from an age distribution that is consistent with the age distribution of the passive smoking cases. The elements c_{01}. and $c_{02}. + c_{03}$. are taken from the study or are estimated from Table B-9. The element c_{02}. is estimated to be 10% of ($c_{02}. + c_{03}$.); c_{03}. is 90%. The elements c_{000} and c_{001} are obtained from c_{00}. and the proportion of never-smoking controls in the study who are married to either never-smokers or ever-smokers. The elements c_{010} and c_{011} are obtained by solving the equations $c_{010} + c_{011} = c_{01}$. and $(c_{000} \times c_{011})/(c_{001} \times c_{010}) = 2.2$. Terms $s_{00} = \sum_{j=1}^{3} c_{0j0}$ and $s_{01} = \sum_{j=1}^{3} c_{0j1}$ are obtained from the equations $s_{00} + s_{01} = s_0$ and $(s_{01} \times c_{000})/(c_{001} \times s_{00}) = 2.8$. Then $c_{020} + c_{030} = s_{00} - c_{010}$ and $c_{021} + c_{031} = s_{01} - c_{011}$. The values of c_{020} and c_{021} are then assumed to be 10% of $c_{020} + c_{030}$ and $c_{021} + c_{031}$, respectively, and c_{030} and c_{031} are assumed to be 90%.

To obtain the elements for the subject cases (i = 1) in Table B-7, it is necessary first to set up relative risks for the passively exposed (k = 1) and not passively exposed (k = 0) wives by observed smoking status (j = 0,1,2,3). These risks are shown in Table B-10.

In most instances, the relative risk, RR(e), for female ever-smokers can be obtained from the study itself or from a related paper (Table B-11). In a few instances, it is necessary to estimate RR(e) from other studies similar in time and locale. In some papers, a current smoker risk also is given. We assume (see explanation above) that the misclassified regular smoker risk, $RR(a)_3$, is equal to the self-reported current smoker risk. Where only RR(e) is available, $RR(a)_3$ can be assumed to be equal to 1.24 × RR(e) based on the data in Table B-12. Because occasional smokers have mean cotinine levels that are 16% of those of regular smokers, it is assumed that $RR(a)_2 - 1 = 0.16(RR(a)_3 - 1)$, and because the former smokers (j = 1) are said to be, on average, long term (Wald et al., 1986; Lee, 1987b), we have averaged the data of Alderson et al. (1985), Lubin et al. (1984), and Garfinkel and Stellman (1988) for the ratio of excess risk of 10+ year former smokers to the excess risk for current smokers and found it to be 9%. Thus, $RR(a)_1 - 1 = 0.09\ (RR(a)_3 - 1)$.

Table B-10. Notation for observed lung cancer relative risks for exposed (k=1) and nonexposed (k=0) wives by the wife's smoking status, using average never-smoking wives $RR(a)_0$ as the reference category

	Wife's smoking status			
Husband's smoking status	Never (j = 0)	Former (j = 1)	Occl. (j = 2)	Reg. (j = 3)
Never (k=0)	RR_{00}	RR_{10}	RR_{20}	RR_{30}
Ever (k=1)	RR_{01}	RR_{11}	RR_{21}	RR_{31}
Weighted avg. active risk	$RR(a)_0 = 1.00$	$RR(a)_1$	$RR(a)_2$	$RR(a)_3$
Passive risk[1]				
$RR(p)_j = RR_{j1}/RR_{j0}$	$RR(p)_0$	$RR(p)_1$	$RR(p)_2$	$RR(p)_3$

[1]Observed passive risk--the ratio of the exposed risk to the unexposed risk in each column.

Table B-11. Prevalences and estimates of lung cancer risk associated with active and passive smoking

	Ever-smokers		Never-smokers		
Case-control	Prev. (%)[1]	Crude RR[2]	Prev. of exposed (%)[3]	Crude RR[2, 4]	Adj. RR[2, 4, 5]
AKIB	21	2.38 (1.67, 3.39)	70	1.52 (0.96, 2.41)	**1.5 (1.0, 2.5)**
BROW[6]	29	4.30 (2.24, 8.24)	15	**1.52 (0.49, 4.79)**	*
			12	1.82 (0.45, 7.36)[7]	1.68 (0.39, 6.90)[7]
BUFF	59	7.06[8] (5.18, 9.63)	84	**0.81[8] (0.39, 1.66)**	*
CHAN	26	3.48 (2.42, 4.99)	47	**0.75 (0.48, 1.19)**	*
CORR	47	12.40 (8.35, 18.4)	46	**2.07[9] (0.94, 4.52)**	*
FONT[10]	43[11]	8.0[11]	63	1.37 (1.10, 1.69)	**1.29 (1.03, 1.62)**
			66	1.21 (0.94, 1.56)	1.28 (0.98, 1.66)
			64	1.32 (1.08, 1.61)	*
GAO	18	2.54 (2.06, 3.12)	74	**1.19 (0.87, 1.63)**	1.34[12,13]
GARF	34[11]	6.0[11]	61	**1.31 (0.93, 1.85)**	1.70[14] (0.98, 2.94)[7]
GENG	41	2.77[15] (1.89, 4.07)	44	**2.16** (1.21, 3.84)	*
HIRA[16]	16	3.20[17] (2.67, 3.83)	77	**1.53[12] (1.10, 2.13)**	1.64[12] *
HUMB	41	16.3 (10.5, 25.1)	56	2.34 (0.96, 5.69)	**2.2 (0.9, 5.5)**
INOU	16	1.66 (0.73, 3.76)	64	**2.55[18] (0.90, 7.20)**	2.54[12,19] *
JANE	42[11]	8.0[11]	68[20]	**0.86 (0.57, 1.29)**	0.93/0.44[21]

(continued on the following page)

Table B-11. (continued)

	Ever-smokers		Never-smokers		
Case-control	Prev. (%)[1]	Crude RR[2]	Prev. of exposed (%)[3]	Crude RR[2, 4]	Adj. RR[2, 4, 5]
KABA[22]	42	5.90 (4.53, 7.69)	60	**0.79 (0.30, 2.04)**	*
KALA	17	3.32 (2.12, 5.22)	60	1.62[23] (0.99, 2.65) 1.41[23] (0.78, 2.55)	**1.92 (1.02, 3.59)[7]**
KATA	28	1.21 (0.50, 2.90)	82	*[24]	*
KOO	32	2.77 (1.96, 3.90)	49	**1.55 (0.98, 2.44)**	1.64
LAMT	24	3.77 (2.96, 4.78)	45	**1.65 (1.22, 2.22)**	*
LAMW	22	4.12 (2.79, 6.08)	56	**2.51[25] (1.49, 4.23)**	*
LEE	60[26]	4.61[26]	68	**1.03 (0.48, 2.20)**	0.75/1.60[27]
LIU	0.05	*	87	0.74 (0.37, 1.48)	**0.77 (0.35, 1.68)**
PERS	37[11]	4.2[11]	43	1.28 (0.82, 1.98)	**1.2 (0.7, 2.1)[7]**
SHIM	21[11]	2.8[11]	56	**1.08[28] (0.70, 1.68)**	*
SOBU	21	2.81 (2.22, 3.57)	54	1.06[23] (0.79, 1.44) 1.77[23] (1.29, 2.43)	1.13[23] (0.78, 1.63)[7] **1.57[23] (1.07, 2.31)[7]**
SVEN	43	5.97 (4.11, 8.67)	66	**1.26[29] (0.65, 2.48)**	1.4[29]
TRIC	10	2.81[30] (1.69, 4.68)	52	**2.08[30] (1.31, 3.29)**	*
WUWI	37	2.24 (1.92, 2.62)	55	**0.79 (0.64, 0.98)**	0.7

(continued on the following page)

Table B-11. (continued)

Case-control	Ever-smokers Prev. (%)[1]	Ever-smokers Crude RR[2]	Never-smokers Prev. of exposed (%)[3]	Never-smokers Crude RR[2, 4]	Adj. RR[2, 4, 5]
BUTL (Coh)	14[11]	4.0[11]	*	2.45[32]	**2.02 (0.48, 8.56)[7]**
GARF (Coh)	22[33]	3.58[33]	72	*	**1.17[12] (0.85, 1.61)[7]**
HIRA (Coh)	16	3.20[17] (1.96, 3.90)	77	**1.38 (1.03, 1.87)**	1.61 *
HOLE[34] (Coh)	56	4.2[11]	73	2.27 (0.40, 12.7)	**1.99 (0.24, 16.7)[7]**

[1]Percentage ever-smokers in controls of whole study (or parent study).
[2]Parentheses contain 90% confidence limits, unless noted otherwise. Crude ORs and their confidence limits were calculated by the reviewers wherever possible. Boldface type indicates values used for analysis in text of this report. OR for case-control studies; relative risk (RR) for cohort studies. The reference category for active smoking is all never-smoking; for passive smoking, it is unexposed never-smokers.
[3]Percentage of never-smoking controls exposed to spousal smoking, unless noted otherwise.
[4]ORs for never-smokers applies to exposure from spousal smoking, unless indicated otherwise.
[5]Calculated by a statistical method that adjusts for other factors (see Table 5-5).
[6]Adenocarcinoma only. Data and OR values communicated from author (Brownson).
[7]95% confidence interval (C.I.).
[8]Exposure to regularly smoking household member. Differs slightly from published value of 0.78, wherein 0.5 was added to all exposure cells.
[9]Excludes bronchioalveolar carcinoma. Crude OR with bronchioalveolar carcinoma included is reported to be 1.77, but raw data for calculation of confidence interval are not provided.
[10]The first, second, and third entries are calculated for population controls, colon cancer controls, and both control groups combined, respectively. For adenocarcinoma alone, the corresponding ORs, both crude and adjusted, are higher by 0.15 to 0.18.
[11]From other studies similar in location and time period (see Table 5-7).
[12]Composite measure formed from categorical data at different exposure levels.
[13]For GAO, data are given as (number of years lived with a smoker, adj. OR): (< 20, 1.0), (20-29, 1.1), (30-39, 1.3), (40+, 1.7).
[14]Estimate for husband smoking 20 cigarettes per day.
[15]Crude OR reported in study is 3.05 (95% C.I. = 1.77, 5.30); adjusted OR is 2.6 (95% C.I. = 1.4, 4.6).
[16]Case-control study nested in the cohort study of Hirayama. OR for ever-smokers is taken from cohort study. This case-control study is not counted in any summary results where HIRA(Coh) is included.
[17]Crude OR is calculated from prospective data in Hirayama (1988). Adjusted OR for ever-smokers given there is 2.67 (no confidence interval).

(continued on the following page)

Table B-11. (continued)

[18]OR reported in study is 2.25, in contrast to the value shown that was reconstructed from the confidence intervals reported in the study; no reply to inquiry addressed to author had been received by press time.
[19]For Inoue, data are given as (number of cig./day smoked by husband, adj. OR): (< 19, 1.58), (20+, 3.09).
[20]Taken from Kabat (1990) as closest in time and place.
[21]From subject responses/from proxy responses.
[22]For second KABA study (see addendum in study description of KABA), preliminary unpublished data and analysis based on ETS exposure in adulthood indicate 68% of never-smokers are exposed and OR = 0.90 (90% C.I. = 0.51, 1.58), not dissimilar from the table entry shown.
[23]For the first value, "ETS exposed" means the spouse smokes; for the second value, "ETS exposed" means a member of the household other than the spouse smokes.
[24]Odds ratio is not defined because number of unexposed subjects is 0 for cases or controls.
[25]Table entry is for exposure to smoking spouse, cohabitants, and/or coworkers; includes lung cancers of all cell types. The OR for spousal smoking alone is for adenocarcinoma only: 2.01 (90% C.I. = 1.20, 3.37).
[26]From Alderson et al. (1985).
[27]From subject responses/from spouse responses.
[28]From crude data estimated to be the following: exposed cases 52, exposed controls 91, unexposed cases 38, unexposed controls 72.
[29]Exposure at home and/or at work.
[30]Known adenocarcinomas and alveolar carcinomas were excluded, but histological diagnosis was not available for many cases. Data are from Trichopoulos et al. (1983).
[31]Raw data for WU is from Table 11 of the Surgeon General's report (U.S. DHHS, 1986). Data apply to adenocarcinoma only.
[32]RR is based on person-years of exposure to spousal smoking. Prevalence in those units is 20%.
[33]Prevalence is calculated from figures in Hammond (1966) for the age distribution of the cases. RR is from U.S. Surgeon General (U.S. DHHS, 1982).
[34]RR values under never-smoker are for lung cancer mortality. For lung cancer incidence, crude RR is 1.51 (90% C.I. = 0.41, 5.48) and adjusted RR is 1.39 (95% C.I. = 0.29, 6.61).

*Data not available.

Table B-12. Observed ratios of current smoker lung cancer risk to ever-smoker risk for females

		Lung cancer RR		Ratio
Study	Exposed cases plus controls	Current smoker	Ever-smoker	Current smoker RR/ ever-smoker RR
Alderson et al. (1985)	901	4.5	4.75	0.95
Buffler et al. (1984)	701	7.9	6.9	1.15
Garfinkel and Stellman (1988)	832	12.7	8.35	1.52
Humble et al. (1985)	268	18.0	13.0	1.38
Svensson et al. (1989)	261	8.46	6.10	1.39
Wu et al. (1985)	317	6.5	4.4	1.48
Overall	3,280	8.05	6.52	1.24[1]

[1]The summary ratio of 1.24 is the geometric mean of the individual ratios weighted by the exposed cases plus controls in that study.

The elements RR_{00} and RR_{01} are obtained from the observed passive relative risk in the study and the never-smoking population weights for controls in Table B-7 by solving the equations

$$1.00 = [(RR_{00} \times c_{000}) + (RR_{01} \times c_{001})]/(c_{000} + c_{001}) \tag{B-3}$$

and

$$RR_{01} / RR_{00} = RR(p)_0. \tag{B-4}$$

Various assumptions regarding passive risks can be used for j = 1,2, and 3. We have assumed, based on the data in Varela (1987), who found that 242 long-term former smokers had essentially the same passive risk as 197 never-smokers, that the passive risk for former smokers is the same as for never-smokers, namely, that $RR(p)_1 = RR(p)_0$. Passive relative risks for female smokers were taken from seven of the passive smoking studies (Akiba et al., 1986; Brownson et al., 1987; Buffler et al., 1984; Humble et al., 1987; Koo et al., 1985; Wu et al., 1985; Hole et al., 1989). The estimates range from 0.7 to 2.3 with no evident trend with either active smoking risk or passive smoking risk. The weighted log mean estimate is 1.25. Since the smokers not exposed to passive smoke already are exposed to considerable ETS from their own smoking, it is probable that the additional ETS from others will have an additive effect rather than a multiplicative effect. Therefore, we have assumed a difference of 0.25 between the active smoking risks of passively exposed and nonexposed current smokers such that $RR_{21} - RR_{20} = RR_{31} - RR_{30} = 0.25$, and $RR_{21}/RR_{20} = RR(p)_2$ and $RR_{31}/RR_{30} = RR(p)_3$. The values for RR_{20} and RR_{30} are derived as follows:

$$RR_{20} = RR(a)_2 - 0.25\ c_{021}/c_{02}., \text{ and } RR_{21} = RR_{20} + 0.25 \tag{B-5}$$

$$RR_{30} = RR(a)_3 - 0.25\ c_{031}/c_{03}., \text{ and } RR_{31} = RR_{30} + 0.25. \tag{B-6}$$

The relative risks for former smokers, RR_{10} and RR_{11}, can be obtained by solving the equations

$$RR(p)_1 = RR_{11}/RR_{10} \tag{B-7}$$

and

$$RR(a)_1 = [(RR_{10}\ c_{010}) + (RR_{11}\ c_{011})]/(c_{010} + c_{011}). \tag{B-8}$$

Crude versions of the elements c_{1jk} (i = 1 for cases) are obtained by multiplying each element c_{0jk} by its respective RR_{jk}. These are then normalized to give the case elements of Table B-7 by

$$c_{1jk} = \frac{c_{0jk}RR_{jk}}{\sum_{j=0}^{3}\sum_{k=0}^{1} c_{0jk}RR_{jk}} \tag{B-9}$$

The next step is to set up Table B-8, which is the table of subjects by observed and true smoking status. This is done by multiplying the observed misclassification rates ($P_{ho}/P_{\cdot j}$) from Tables B-3 and B-4 by the appropriate elements from Table B-7. For example, $P_{10} = c_{01}.(P_{10}/P_{\cdot 1})$. An attempt was made to use the true misclassification rates from Tables B-3 and B-4 on the theory that they would exhibit less variability in being transferred from the cotinine and discordant answer studies to the passive smoking calculations. However, the method is laborious and, as is shown in the Correa example below, does not lead to increased accuracy.

The next step is to develop a deletions table to implement Equation B-1 above using the control and case smoking prevalences in Table B-7 and the distribution in Table B-8. Each observed element, c_{i0k}, in Table B-7 is multiplied by its appropriate observed misclassification factor, $P_{h0}/P_{\cdot j}$, where h = j, to yield a deletion element to be subtracted from the appropriate observed wives' never-smoking-status elements: c_{000}, c_{001}, c_{100}, and c_{101}, to obtain corrected elements C_{000}, C_{001}, C_{100}, and C_{101}. Thus,

$$C_{000} = c_{000} - \sum_{h=j=1}^{3} c_{0j0}\ P_{h0}/P_{\cdot j}, \text{ etc.} \tag{B-10}$$

Once these corrected never-smoker elements are obtained, the relative risk corrected for smoker misclassification is obtained from Equation B-2; $RR(c)_0 = (C_{101} \times C_{000})/(C_{100} \times C_{001})$, and the bias becomes $RR(p)_0 / RR(c)_0$.

B.6. NUMERICAL EXAMPLE

Using the Correa et al. (1983) study as an example, the study tells us that 52.8% of the wives never smoked and that 45.9% of the never-smoking wives were exposed to their spouses' smoke. This establishes $c_{00}.$ as 0.528 and c_{000} and c_{001} as 0.286 and 0.242, respectively. The quantity $s_0.$, the proportion of ever-smokers, by difference is 0.472. From Correa's Table 2 we find that the former smokers are 35.5% of the ever-smokers. Thus, the former smokers, $c_{01}.$, become 0.167, and the current smokers ($c_{02}. + c_{03}.$) become 0.305. The current smokers are divided into current regular smokers at 90% ($c_{03}. = 0.275$) and current occasional smokers at 10% ($c_{02}. = 0.030$). These data are shown in the bottom line of Table B-13.

Using the concordance factor of 2.8 for ever-smokers versus never-smokers, it is possible to show as described above that 33.2% of the females in the Correa study would be ever-smoker wives with smoking husbands (s_{01}) and that 14.0% would be ever-smoker wives with never-smoking husbands (s_{00}). Similarly, using the concordance factor of 2.2 for former smoking wives and ever-smoking husbands versus the never-smokers, the former smoking wives married to ever-smoking husbands (c_{011}) would be 10.9% of the total and those married to the never-smoking husbands (c_{010}) would be 5.8%. Then by difference, exposed current smoking wives ($c_{021} + c_{031}$) would be 22.3%, to be split into 20.1% regular smokers (c_{031}) and 2.2% occasional smokers (c_{021}), and the nonexposed current smoking wives ($c_{020} + c_{030}$) would be 8.2%, split into 7.4% regular smokers (c_{030}) and 0.8% occasional smokers (c_{020}). These data now supply all the elements needed in Table B-13 and the control part of Table B-7.

The estimate of relative risk for passive smoking, $RR(p)_o$, for females is 2.07 (Correa et al., 1983). The age- and sex-adjusted relative risk for current smoking from a related paper

Table B-13. Observed smoking prevalence among the controls--Correa example

Husband's smoking status	Wife's smoking status: Never	Former	Occasional	Regular	All
Never	0.286	0.058	0.008	0.074	0.426
Ever	0.242	0.109	0.022	0.201	0.574
All	0.528	0.167	0.030	0.275	1.000

(Correa et al., 1984) is 12.6. The ratio of female smoking crude risk to the average for males and females is about 80%, indicating an age-adjusted current female risk of about 10. (*Note*: This is different from the current smoker relative risk that would be calculated from the crude ever-smoker risk of 12.4 used in Table 5-7 [of this report] and Table B-3. The adjusted risk is used here simply as an example.) With these inputs and the weights of controls in the study, the various exposed and nonexposed relative risks are those shown in Table B-14. The weighted average risk for the occasional smokers is calculated as 0.16 (current regular risk − 1) + 1, which for this example is 0.16 (10 − 1) + 1 = 2.44. The weighted average risk for former smokers is 0.09 (current regular risk − 1) + 1, which is 0.09 (10 − 1) + 1 = 1.81. The weighted average risks are split between never-smoking and ever-smoking husbands by using the passive risks, the population weights, and Equations B-3, B-4, B-5, B-6, B-7, and B-8. A crude case prevalence table is then made up (Table B-15) by multiplying each c_{0jk} by its respective RR_{jk}. This table is then normalized (Equation B-9) by dividing by 3.653 to yield Table B-16, which is the lower half of Table B-7 for this example.

The smoking status distribution table (Table B-17) is developed, as described above, from the misclassification factors in Tables B-3 and B-4 and the bottom line of Table B-13. For example, to arrive at element (h = 3, j = 0), the observed $P_{\cdot 3}$ of 0.275 is multiplied by an observed misclassification factor of 0.0109 (from Table B-3) to yield 0.003. To explore the value of using the true misclassification factors instead of the observed ones, the true and observed m's were carried to five decimal places. An approximation procedure to determine the true smoking probabilities $P_{0\cdot}$, $P_{1\cdot}$, $P_{2\cdot}$, and $P_{3\cdot}$ was carried through four stages. The resulting total true distribution of smoking status rounded to three decimal places was essentially identical to the distribution shown in the bottom line of Table B-17. Similarly, any differences in the individual elements were very small and beyond the accuracy of the underlying data. The Correa study was

Table B-14. Observed relative risks--Correa example

Husband's smoking status	Wife's smoking status Never (j = 0)	Former (j = 1)	Occasional (j = 2)	Regular (j = 3)
Never	0.67	1.07	2.26	9.82
Ever	1.39	2.21	2.51	10.07
Weighted average	1.00	1.81	2.44	10.00
Passive risk, RR(p),	2.07	2.07	1.11	1.025

Table B-15. Crude case table, prevalence of cases by smoking status--Correa example

Husband's smoking status	Wife's smoking status Never	Former	Occasional	Regular	All
Never	0.192	0.062	0.018	0.726	0.998
Ever	0.336	0.240	0.055	2.024	2.655
All	0.528	0.302	0.073	2.750	3.653

Table B-16. Normalized case table, prevalence of cases by smoking status--Correa example

Husband's smoking status	Never	Wife's smoking status Former	Occasional	Regular	All
Never	0.052	0.017	0.005	0.199	0.273
Ever	0.092	0.066	0.015	0.544	0.727
All	0.144	0.083	0.020	0.743	1.000

Table B-17. Distribution of subjects by observed and true smoking status for wives in Correa example[1]

Wife's observed smoking status	Never (h = 0)	Wife's true smoking status Former (h = 1)	Occasional (h = 2)	Regular (h = 3)	All
Never (j = 0)	0.499	0.019	0.007	0.003	0.528
Ex (j = 1)	0	0.160	0.004	0.003	0.167
Occasional (j = 2)	0	0	0.030	0	0.030
Regular (j = 3)	0	0	0	0.275	0.275
All	0.499	0.179	0.041	0.281	1.000

[1] Values rounded to three decimal places.

chosen as our example because the female ever-smoking prevalence is reasonably high (47.2%) and the female current smoker lung cancer relative risk is high (10), both of which are factors that should lead to a greater rather than a smaller correction to the passive risk.

We now can set up a deletions table, Table B-6, which is the equivalent of equations B-1 and B-10 above, by multiplying the control and case elements in Table B-13 and B-16 by the appropriate observed misclassification rates $P_{h0}/P_{\cdot j}$ (h = j), namely, $P_{10}/P_{\cdot 1} = 0.117$, $P_{20}/P_{\cdot 2} = 0.242$, and $P_{30}/P_{\cdot 3} = 0.0109$. For example, to get 0.00679, one multiplies 0.058 from Table B-10 by 0.117. Then the first three columns are summed horizontally to get the fourth column, which is then subtracted from the elements in the "never" columns of Tables B-13 and B-16 (column 5) to get the "corrected never" elements (column 6).

The corrected passive risk is now obtained by taking the cross-product from the "corrected never" column: (0.07461 × 0.27647)/(0.04666 × 0.22173) = 1.99, which is to be compared with the observed risk of 2.07. The bias is then 2.07/1.99 = 1.04. It is interesting to note how sensitive the bias is to the smoker relative risk that is assumed. When the crude smoker risk (no age adjustment) of 12.4 for ever-smokers, equivalent to about 15.4 for current regular smokers, is assumed, the corrected passive risk estimate is 1.89 and the bias is twice as great at 1.10.